基于微博大数据分析的未成年期遭受性侵对女性受害者心理影响的研究

万文雨[1 2]孙玖陌[1 2] 刘洁含[1 2] 杨舒雯[1 2] 朱廷劭[2*]

[1]（中国科学院大学 北京 100049）

[2]（中国科学院心理研究所 北京 100101）

摘要：

[**背景**]童年性经历对一个人身心的不利影响是持久的。然而，很少有研究关注这一群体的心理健康状况。[**目的**]本研究的目的是基于微博大数据心理计算模型来探究儿童期有性虐待经历的成年人（CSA 组）与无此经历的成年女性（对照组）在社会态度、幸福感、心理健康在内的心理特征的差异。[**方法**]本研究在新浪微博上收集了 46 名受害者（全部是女性）和 46 名非受害者（与 CSA 组别的性别匹配），爬取所选用户的全部微博，采用微博大数据计算模型来计算其在各项心理特征的得分。[**结果**]采用独立样本 t 检验，研究结果显示，在社会态度、幸福感、特别是心理健康方面，两者存在显著差异，与此同时我们还发现两组在微博行为特征上存在差异。与非受害者相比，受害者在抑郁、压力等健康特征得分较高，在生活满意度、自我接纳等心理特征得分较低，但都未达到心理疾病诊断等阈限临界值。[**局限**]基于微博大数据模型与用心理测量量表测的的社会态度、幸福感、心理健康水平无法完全等同，不同完全替代严格的心理测量。[**结论**]童年期性情经历对成年女性对心理特征有负面影响，但是这种影响不足以达到患心理疾病的阈限标准。

关键词：童年期性侵；社交媒体；心理特征；大数据

分类号：TP393

Using social media to explore the psychological features of the female adults with childhood sexual abuse

Wenyu Wan[1,2], Jiumo Sun[1,2], Jiehan Liu[1,2], Shu-Wen Yang[1,2], Tingshao Zhu[1*]

[1](Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China)

[2](University of Chinese Academy of Sciences, Beijing, 101407, China)

Abstract:

[Background] The adverse effects of childhood sexual abuse experience on female physical and psychological health are enduring. However, few studies have focused on the psychological characteristics of this group when they grew up.

[Objective] The purpose of this study was to explore the difference in psychological characteristics including social attitudes, well-being, and mental health between the females with sexual abuse experience in childhood (CSA group) and females without this experience (control group) based on the microblogging data calculation model.

[Methods] This study collected 46 victims (all females) and 46 non-victims (sex matching with CSA group) on Sina Weibo, crawled all the microblogs of the selected users and and calculate its

score on various psychological characteristics by microblogging data calculation model.
[Results] Using independent sample t-test, the results showed that there were significant differences in social attitudes, well-being, and especially mental health. At the same time, we also found that there were differences in microblog behavior characteristics between the two groups. Compared with non-victims, the victims had higher scores in depression, stress and other health characteristics, and lower scores in psychological characteristics such as life satisfaction and self-acceptance. However, they did not reach the critical thresholds for the diagnosis of mental diseases.
[Limitations] The psychological features obtained from microblogging data calculation model can not completed equivalent to the psychological features obtained from psychological scales and can not replace the rigid psychological measurement.
[Conclusion] The childhood sexual abuse experience has negative effect on female. However, this effect is not sufficient to meet the threshold criteria for mental illness.
Keywords: Childhood sexual abuse, Social media, Psychological characteristics, Big data

1 引言

未成年期性侵害是一个世界性的问题。一项针对 21 个国家（其中大多为发达国家）的研究指出，有 7%~36%的女性和 3%~29%的男性报告在未成年阶段曾经遭受过性侵害[1]。目前对性侵害行为的最广泛分类主要包括性侵入、性接触及没有接触的性三类[2]。性侵害对受害者的影响是多方面的，如愤怒、焦虑、抑郁、恐惧、孤独、低自尊、创伤后应激障碍（PTSD）、自杀倾向等心理问题；人际关系困难、学业成绩差、信任困难等社会交际问题；性早熟、性攻击等性问题[2]。性侵害经历对受害儿童的身心伤害是严重而持久的，极有可能延伸至成年阶段。

随着 Web 2.0 时代的到来，在线社交媒体凭借其开放性、交互性和实时性等特点获得了空前的发展。微博作为国内最具代表性的在线社交媒体之一，截止 2017 年 5 月，月活跃用户已突破 3.4 亿，超过 Twitter 首次跃居世界第一。通过微博平台，人们可以进行在线互动，并自由公开地对时事热点表达自己的观点和情感。人类丰富的心理过程藉由微博上的公开表达得以显现，使得微博成为心理学等社会科学研究的绝佳工具[3]。与此同时，大数据研究方法的蓬勃发展也为进行在线文本分析提供了方法论上的支持。

2017 年下半年，有关性侵害的新闻报道层出不穷："红黄蓝"幼儿园涉嫌存在性侵幼童行为、北电教授侯亮平性侵女学生、江西南昌大学被曝存在教师性侵学生事件、北航毕业女博士罗茜茜实名举报长江学者陈小武性骚扰女学生……"性侵"这一社会还未曾根除的毒瘤再一次进入大众的视野，每一件新闻都在微博平台受到了民众的激烈讨论。由于微博用户的隐私性，越来越多曾经的性侵受害者在新闻微博下评论，鼓起勇气说出自己曾经遭受的伤害，甚至发布原创微博，希望以此警醒社会，性侵就发生在我们身边，受害者的年龄已经趋向低龄化。这些评论和原创微博都对本研究课题提供了大量丰富的数据来源，使得本研究得以通过微博大数据的心理模型研究童年的性侵害经历对女性受害者的心理影响。

我国教育部在 2007 年 2 月发布的《中小学公共安全教育指导纲要》中提出，未成年中小学生应该了解应对性侵害的一般方法和技能[4]。对未成年性侵害问题进行研究可以为针对性教育提供资料。但目前国内关于这方面的研究很少且实验大多使用同一专业的大学生这样的单一样本，对性侵害问题的研究还存在很多不足[5-7]。通过社交网络进行被试的选取克服了样本选取单一的限制，可以为国家

的教育、卫生事业提供更丰富的资料。

早期对于性侵害经历对受害者的心理健康影响的研究多采用问卷调查的方法，这诚然可以做到对受害者相关情绪与行为特征的精准掌握，但仍存在以下问题。首先，目前国内尚无标准的儿童期性侵害问卷，此前的研究多为自编，难免有所疏漏。其次，虽然大多数问卷调查研究在实施过程中注重隐私，但被试仍会害怕暴露隐私而选择不参加实验或在问卷填写中回避真实信息，尤其问卷中所陈述的问题大多非常直白，很容易使受害者回想起曾经的痛苦经历，极易给被害者造成二次伤害。除此之外，问卷法费时费力，且收集到的数据量往往无法达到研究要求。而本研究的数据来源均为曾在微博平台上提及自己是性侵受害者的女性微博用户的原创微博，首先保证了样本量；其次，由于微博用户均为匿名，且被提取的用户并不知情，因此本研究的隐私性也得到了保证；最后，我们运用已经成熟的基于微博大数据分析的心理模型对用户的原创微博进行分析，分析结果保证了客观性、标准性。

基于以上论述，本研究通过心理模型分析受害者的原创微博，了解女性童年遭受性侵害与成年后心理健康之间的关系，分析女性受害者成年后在社会态度、幸福感、心理健康方面等方面是否与控制组存在差异。

2 方法

2.1 被试选取

本研究的对象是从微博中选取的在未成年时期有过性侵经历的成年女性用户（实验组）以及没有该经历的成年女性用户（控制组）。对于实验组被试的选取需满足以下 4 个条件：1）微博注册为女性用户；2）该用户现已成年；3）在他人微博下评论或自己原创微博中明确带有“自己小时候被性侵”的字眼；4）用户为私人普通用户，排除公共人物账号、公共账号、运营号和僵尸号。对于控制组被试的选取需满足以上第 1）、2）、4）条条件以外，还应在个人微博中无任何叙述自己遭受过性侵的微博或评论。最终选取实验组 46 人，控制组 46 人。

2.2 数据收集与处理

人工在热门有关性侵的微博下按照以上标准删选评论，记录用户的 ID，通过新浪微博 API（Application Programming Interface，应用程序编程接口）爬取每个用户一些基本信息（性别、VIP 认证信息、粉丝数、关注数等）以及所有微博。由于转发微博的内容并非用户的自我表达，因此在分析中剔除转发微博，只保留原创微博和转发微博中的原创部分。本研究采用基于微博数据的心理模型对微博用户的所有原创内容进行社会态度方面（生活满意度、收入满意度、社会地位、国家政府满意度、地方政府满意度、国家政府信心、地方政府信心、国家经济满意度、地方经济满意度、社会风险判断、社会公平满意度、愤怒情绪、集群效能、集群行为意向）、幸福感方面（环境控制、与他人积极关系、消极情绪、Oxford-Happiness、自主性、个人成长、自我接纳、生活满意度、生活目标）、以及心理健康（焦虑、抑郁、压力等）方面的计算。其中，在衡量心理健康方面，采用抑郁焦虑压力问卷（DASS，分为焦虑、抑郁、压力三个子维度得分）和症状自评量表（SCL-90，分为躯体化、焦虑、强迫症状、抑郁、人际关系敏感性、精神病性、偏执、敌对、恐怖）。随后，对数据进行独立样本 t 检验探究实验组和控制组是否在社会态度、幸福感、心理健康方面存在差异。

3 结果

表 1 受害者与控制组微博基本情况差异分析

维度	组别	人数	平均数	标准差	t	p
是否认证	受害者	46	1.978	1.474	0.00	1.000
	控制组	46	1.978	1.474		
性别	受害者	46	2.000	0		
	控制组	46	2.000	0		
粉丝数	受害者	46	550.783	1737.433	-0.50	0.617
	控制组	46	688.565	671.111		
关注数	受害者	46	377.717	523.896	-0.54	0.588
	控制组	46	430.935	409.140		
互粉数	受害者	46	16.891	56.008	-4.220	0.000
	控制组	46	138.978	188.053		
微博数	受害者	46	1105.913	1735.159	-3.775	0.000
	控制组	46	2646.304	2155.783		

对受害者和控制组的微博基本情况进行独立样本 t 检验（df=90），结果表明两组被试的互粉数与微博数差异显著（$t_{互粉数}$=-4.220，$p_{互粉数}$=0.000<0.001，d=0.880；$t_{微博数}$=-3.775，$p_{微博数}$=0.000<0.001，d=0.787），受害者的互粉数量显著少于控制组，且受害者的微博数量也同样显著少于控制组。

表 2 受害者与控制组社会态度差异分析

维度	组别	人数	平均数	标准差	t	p
生活满意度	受害者	46	3.428	0.194	-1.735	0.088
	控制组	46	3.481	0.076		
收入满意度	受害者	46	2.509	0.283	1.936	0.056
	控制组	46	2.419	0.136		
社会地位满意度	受害者	46	2.645	0.312	2.551	0.012
	控制组	46	2.518	0.129		
国家政府满意度	受害者	46	2.929	0.239	-1.315	0.194
	控制组	46	2.978	0.083		
地方政府满意度	受害者	46	2.516	0.310	0.209	0.835
	控制组	46	2.505	0.149		
国家政府信心	受害者	46	2.680	0.254	-1.158	0.251
	控制组	46	2.729	0.130		
地方政府信心	受害者	46	2.703	0.173	1.742	0.086
	控制组	46	2.654	0.079		
国家经济满意度	受害者	46	3.259	0.141	-1.217	0.227
	控制组	46	3.293	0.125		
地方经济满意度	受害者	46	3.106	0.198	-1.147	0.255
	控制组	46	3.145	0.118		
社会风险判断	受害者	46	4.119	0.163	1.640	0.106
	控制组	46	4.075	0.081		
社会公平满意度	受害者	46	3.078	0.421	1.775	0.082
	控制组	46	2.964	0.119		

愤怒情绪	受害者	46	1.637	0.328	0.027	0.979
	控制组	46	1.635	0.086		
集群效能	受害者	46	3.522	0.248	0.523	0.603
	控制组	46	3.501	0.093		
集群行为意向	受害者	46	2.586	0.306	0.379	0.706
	控制组	46	2.567	0.120		

对受害者与控制组的社会态度进行独立样本 t 检验(df=90)，$t_{\text{社会地位满意度}}=2.551$，$p_{\text{社会地位满意度}}=0.012<0.05$，d=0.532，受害者对自身的社会地位的满意程度显著高于控制组，而在其他方面的社会态度上，两组没有显著差异。

表 3 受害者与控制组 DASS 量表得分差异分析

维度	组别	人数	平均数	标准差	t	p
焦虑	受害者	46	4.113	0.629	1.066	0.290
	控制组	46	3.983	0.525		
抑郁	受害者	46	5.269	1.069	2.087	0.040
	控制组	46	4.883	0.660		
压力	受害者	46	3.522	1.236	2.617	0.011
	控制组	46	3.018	0.418		

对受害者和控制组在 DASS 量表上的得分进行独立样本 t 检验（df=90），结果表明，受害者在抑郁和压力水平上显著高于控制组($t_{\text{抑郁}}=2.087$，$p_{\text{抑郁}}=0.040<0.05$，d=0.435；$t_{\text{压力}}=2.671$，$p_{\text{压力}}=0.011<0.05$，d=0.546)，而两组的焦虑水平没有显著差异。

表 4 受害者与控制组幸福感得分差异分析

维度	组别	人数	平均数	标准差	t	p
环境控制	受害者	46	11.167	0.563	0.073	0.942
	控制组	46	11.160	0.223		
与他人的积极关系	受害者	46	11.994	0.497	-2.217	0.029
	控制组	46	12.203	0.406		
积极情绪	受害者	46	22.225	1.209	-1.024	0.059
	控制组	46	22.610	0.611		
消极情绪	受害者	46	15.636	1.151	1.729	0.087
	控制组	46	15.277	0.810		
Oxford Happiness	受害者	46	85.818	3.580	-2.720	0.008
	控制组	46	87.591	2.600		
自主性	受害者	46	10.292	1.213	-1.658	0.103
	控制组	46	10.604	0.400		
个人成长	受害者	46	14.486	1.007	0.976	0.333
	控制组	46	14.330	0.414		
自我接纳	受害者	46	9.590	0.822	-2.176	0.034
	控制组	46	9.872	0.309		
生活满意度	受害者	46	13.402	1.658	-2.196	0.032
	控制组	46	13.970	0.576		
生活目标	受害者	46	11.977	0.937	-0.287	0.775
	控制组	46	12.021	0.437		

幸福感方面，独立样本 t 检验（df=90）表明，受害人组在与他人的积极关系，自我接纳和生活满意度方面，得分显著低于控制组（$t_{与他人的积极关系}$=-2.217，$p_{与他人的积极关系}$=0.029<0.05，d=0.461； $t_{自我接纳}$=-2.176，$p_{自我接纳}$=0.034<0.05，d=0.454；$t_{生活满意度}$=-2.196，$p_{生活满意度}$=0.032<0.05，d=0.458）；而 Oxford Happiness 得分非常显著低于控制组（$t_{Oxford\ Happiness}$=-2.720，$p_{Oxford\ Happiness}$=0.008<0.01，d=0.567）

表 5 受害者与控制组 SCL-90 量表得分差异分析

维度	组别	人数	平均数	标准差	t	p
躯体化	受害者	46	0.757	0.208	2.394	0.020
	控制组	46	0.676	0.098		
强迫	受害者	46	1.402	0.241	4.742	0.000
	控制组	46	1.215	0.117		
人际关系敏感	受害者	46	1.302	0.264	4.638	0.000
	控制组	46	1.098	0.139		
抑郁	受害者	46	1.115	0.268	3.914	0.000
	控制组	46	0.985	0.135		
焦虑	受害者	46	0.411	0.063	1.066	0.290
	控制组	46	0.398	0.053		
敌对	受害者	46	1.128	0.325	2.227	0.029
	控制组	46	1.010	0.158		
恐怖	受害者	46	0.813	0.254	3.564	0.001
	控制组	46	0.658	0.148		
偏执	受害者	46	1.062	0.195	3.814	0.000
	控制组	46	0.941	0.094		
精神病性	受害者	46	0.985	0.228	2.906	0.005
	控制组	46	0.879	0.098		

对于受害者和控制组的 SCL-90 量表得分进行独立样本 t 检验（df=90），除焦虑分量表得分没有显著差异外，受害者在躯体化，敌对分量表上的得分显著高于控制组（$t_{躯体化}$=2.394，$p_{躯体化}$=0.020<0.05，d=0.498；$t_{敌对}$=2.227，$p_{敌对}$=0.029<0.05，d=0.461）；受害者在恐怖和精神病性分量表的得分非常显著地高于控制组（$t_{恐怖}$=3.564，$p_{恐怖}$=0.001<0.01，d=0.746；$t_{精神病性}$=2.906，$p_{精神病性}$=0.005<0.01，d=0.604）；而在强迫，人际关系敏感，抑郁，偏执等分量表上的得分，受害者与控制组达到了极其显著水平。（$t_{强迫}$=4.742，$p_{强迫}$<0.000，d=0.987；$t_{人际关系敏感}$=4.638，$p_{人际关系敏感}$<0.000，d=0.967；$t_{抑郁}$=3.914，$p_{抑郁}$<0.000，d=0.613；$t_{偏执}$=3.814，$p_{偏执}$<0.000，d=0.790）

4 讨论

本研究旨在借助基于微博大数据的心理计算模型来探究儿童期有性虐待经历的成年人与无此经历的成年女性在包括社会态度、幸福感、心理健康在内的心理特征的差异。研究结果显示，在社会态度、幸福感、特别是心理健康方面，两者存在显著差异，与此同时我们还发现两组在微博行为特征上存在差异。

在社会态度方面，本研究在生活满意度、收入满意度、社会地位满意度等 14 个方面来探究实验组和控制组之间的差异，发现在收入满意度和社会地位满意度上两组存在显著差异其他维度均不显著，具体来说，实验组在这两个维度的

得分高于控制组。出现此结果，一方面可能是由于有创伤经历的人，在之后的生活中更加积极的工作、学习，来获得更高的收入和社会地位，以努力摆脱创伤经历的不利影响；另一方面可能是由于有性侵创伤经历的人对于自己的收入预期和社会地位预期相较于控制组偏低，因此对于现有的收入状况和社会地位及其容易满足。

在幸福感方面，本研究在在环境控制、与他人的积极关系、积极情绪、消极情绪等 10 个方面来探究两组之间的差异，发现在与他人积极关系、积极情绪、Oxford-Happiness、自我接纳、生活满意度这五个方面两组存在显著差异，且结果均是实验组低于控制组，与实验预期结果相同。其实不难想象，性侵经历是创伤的一种，特别是在我国的传统观念内，以有过性侵经历为耻，更有甚至由于自身以及家人受到传统观念的束缚，即使有此遭遇，也不愿主动报警，使得犯罪者逍遥法外，这无疑对受害者来说又是一大伤害。而这种伤害无疑是持久的，伴随一生的，影响了自己与他人的人际交往关系、自我接纳、幸福感等。值得注意的是，在幸福感维度下的生活满意度实验组显著低于控制组，而在社会态度下的幸福感实验组和控制组的差异并不显著，这可能是计算机模型不同而导致对微博文本特征提取不明感导致的。需要进一步通过更优化的计算机模型来验证。

在心理健康方面，本研究主要想探究实验组在抑郁、焦虑等心理疾病上的情况相较于控制组是怎样的，选取了 DASS 和 SCL-90 两个计算机模型，其中 DASS 包括焦虑、抑郁和压力 3 个方面，SCL-90 包括躯体化、焦虑、强迫症状等 9 个方面。研究结果表明，DASS 中，实验组在抑郁和压力两维度的得分显著高于控制组，但在焦虑维度上不显著；SCL-90 中，实验组在躯体化、焦虑、强迫症状、抑郁、人际关系敏感、精神病性、偏执、敌对、恐怖这 9 个方面均高于控制组。不难看出两者有在探究被试的抑郁、焦虑方面存在重叠，而研究结果在两个问卷中结果是一致的，即在抑郁维度上显著，在焦虑维度上两者的差异不显著。值得说明的一点，虽然实验组在这些维度上的得分显著高于控制组，但是其得分还是在标准线以下，即没有到达病症的底部，尚属于健康人群。但这同时给我们一些信号，有性侵经历的人在抑郁、压力、偏执等方面的得分高于健康，我们在平时应注重对该人群的心理辅导，通过多种干预手段来提高该群体的心理健康水平。

本研究存在一定的局限性。首先，在样本选取方面可能与总体仍旧存在一定的偏差，样本只能代表未成年期具有性侵经历的成年女性微博用户的群体，而不能很好的代表未成年期具有性侵经历的成年女性。究其原因，在微博上明确表述自己有性侵经历的人可能相较于那些没有微博的或是没在微博上表述自己有相似经历的人，本身在社会态度、幸福感、心理健康等方面就具有巨大的差异。其次，在研究方法层面上，微博短文文本特征关键词数量少，可能在一定程度上会妨碍特征计算，导致计算结果的不准确。

尽管存在诸多局限性，本研究率先基于微博大数据模型通过分析两个群体在社交平台发布的文字信息来获取多种心理特征，并指出这两群体在语言表达上的差异，克服了对被试二次伤害以及被试难找的问题。有研究显示，童年期有性侵经历的人相较于无性侵经历的人，他们自杀倾向、患精神疾病如抑郁症、焦虑症的风险更高[8-10]。而自杀倾向、或患精神疾病的趋向性往往从他们的心理特征表现出来。近期，一项研究通过监控个体在社交媒体上的语言表达来对其进行自杀干预[11]。这为我们提供了一个很好的研究思路。童年期有性侵经历的人作为自杀的高发人群，对其发布的文字进行密切的关注，以便及时对其进行自杀干预。此外，本研究方法可以作为对该人群干预效果治疗的一个有效评估方法。

参考文献

[1] PINHEIRO P S. World Report on Violence against Children [M]. 2006.

[2] 李成齐. 性侵害受害儿童的心理病理学研究 [J]. 中国特殊教育, 2007, 02): 71-5+59.

[3] 乐国安，董颖红，陈浩，et al. 在线文本情感分析技术及应用 [J]. 心理科学进展，2013，10): 1711-9.

[4] 教育部. 中小学公共安全教育指导纲要 [N]. 2007-03-01.

[5] 陈晶琦，韩萍，P.DUNNE M. 892 名卫校女生儿童期性虐待经历及其对心理健康的影响 [J]. 中华儿科杂志, 2004, 01): 41-5.

[6] 谷来有，迟淑艳，张兆华. 某高校大学生儿童期性虐待情况 [J]. 中国学校卫生, 2005, 04): 309-10.

[7] 孙言平，董兆举，衣明纪, et al. 1307 名成年学生儿童期性虐待发生情况及其症状自评量表测试结果分析 [J]. 中华儿科杂志, 2006, 01): 21-5.

[8] EGGERTSON L. High rates of childhood abuse, depression in Inuit suicides [J]. Canadian Medical Association Journal, 2013, 185(10): E433-E4.

[9] LOPEZ-CASTROMAN J, MELHEM N, BIRMAHER B, et al. Early childhood sexual abuse increases suicidal intent [J]. World Psychiatry, 2013, 12(2): 149-54.

[10] CHOI K W, BATCHELDER A W, BOROUGHS M, et al. EXAMINING THE SYMPTOM NETWORK OF DEPRESSION AND PTSD IN RELATION TO SEXUAL RISK BEHAVIOR IN MSM WITH CHILDHOOD SEXUAL ABUSE [J]. Annals of Behavioral Medicine, 2017, 51(S1779-S80.

[11] TAN Z, LIU X, LIU X, et al. Designing Microblog Direct Messages to Engage Social Media Users With Suicide Ideation: Interview and Survey Study on Weibo [J]. Journal of Medical Internet Research, 2017, 19(12): e381.

(通讯作者:朱廷劭 E-mail:tszhu@psych.ac.cn)

作者贡献声明：

万文雨：提出研究思路、设计研究方案、撰写摘要、方法和讨论部分

孙玖陌：提出研究思路、设计研究方案、撰写引言部分

刘洁含：提出研究思路、设计研究方案、撰写引言部分

杨舒雯：提出研究思路、设计研究方案、数据分析、撰写结果部分、格式修改

朱廷劭：研究思路指导、提供数据样本、论文指导